MATHEMATICAL-STATISTICAL INVESTIGATIONS ON THE PIGMENT CONTENT OF SZEGED PAPRIKA MILLING PRODUCTS IN THE YEARS 1959 TO 1967

By

T. HUSZKA[1]

School of Food Technology, Szeged

(Received November 25, 1968)

One of the most important qualitative characteristics of the milling products of paprika: the total pigment content and its changes during the period 1959 to 1967 have been studied. The usual statistical characteristics (x, s, R, V) were determined from 7100 analytical data of 5 milling products of different qualities.

On the basis of KRAMER's "non parametric" test the deviations between the pigment contents of milling products of different qualities are shown to be significant in the period of study, whereas the deviations within the same quality in different years were not significant.

In an earlier paper [1] it was pointed out that the processing of the milling products of paprika (Capsicum annuum L. var. longum Szegediense) can be optimized economically by using linear programming for calculating the quantities of the half products available to be taken in order to obtain a product which fulfils the specifications for the quality in question. The condition of using linear programming is, however, that the qualitative components determining the quality both of the end products and of the primary material can be expressed by numerical characteristics.

In this paper the pigment content of the milling products of paprika is studied as a numerical qualitative characteristic. It must be pointed out that, — though Hungarian paprika has been marketed after obligatory state-control for more than fifty years now, this ensuring its constant good quality — still there do not exist prescribed numerical characteristics for each property of the products.

Milling products of different qualities (exquisite-delicacy, noble-sweet, half-sweet, rose, hot) have numerical characteristics, expressed in %, for their ash, sand, oil content, etc. specified in the respective standards of qualification [2], but there is no inland prescription for the evaluation of the total pigment content. Concerning its colour, paprika produced for the home market is assessed by comparing its colour visually with that of a standard sample, accepted jointly by the representatives of the producers, commerce and OMMI[2]. The examination has to ascertain

[1] From the Ph. D. theses submitted to the Attila József University Szeged (1968).
[2] State Institute for Qualification of Agricultural Products

whether the colour of the product in question is the same as that of the respective standard sample, failing which the product is declared as unsatisfactory. This method does not yield any numerical results for the pigment content.

In the foreign trade there are, however, prescriptions on the part of several countries not only as to the colours matching the samples, but also concerning the determination of the total pigment content; therefore also this value has to be determined in the course of the official qualification performed by the OMMI and recorded in the quality certificate issued. The determination of pigment content is carried out with BENEDEK's [3] photometrical method, and the results are given as g/kg pigment content expressed in capsantin. The pigment content determined in this way can easily be converted [4] into ASTA[3] units used in the USA [5].

Thus, it seemed suitable to consider the results of the measurements of the pigment contents, listed on the quality certificates of OMMI, as a basis for determining the numerical characteristics of the pigment content in the course of the processing of programmed milling products.

The numerical measurement of the total pigment content is to be considered as a very important information concerning the quality of the milling product, because its value is determined by the carotin-like pigment content of the pericarpium [6] and this parameter is measurable in every stage of the processing. It must be born in mind that both manufacturing and qualifying of paprika milling products is concerned with a biological substance, the compositon of which depends on several factors (weather, soil etc.) too; besides, the pigment content values of different batches in processing the same quality may also differ in many cases. Therefore the results of measurements of the pigment content can only be suitably studied with mathematical-statistical methods, based on data obtained from a great number of samples. It is to be expected, however, that the mathematical-statistical analysis will result not only in determining the values of the usual characteristics, but it may also yield information on some controversial questions, such as the necessity of establishing new standard samples for every year, and also the expediency of the qualification on the basis of the total pigment content, instead of standard samples, in quality-controls.

Methods of Statistical Analysis and Evaluation of Data

The data subject to statistical analysis were collected from the results of measurements of the pigment content of export milling products of five different qualities, carried out at the Department of Paprika Control of OMMI in the years 1959 to 1967. The number of the analysed data was about 7100. Within each quality, the pigment content of the analysed samples can be considered to be a random variable, more exactly, a continuous random variable. The data collected and grouped according to years and qualities have been shown in histograms. Considering the accuracy of measurement of the Pulfrich photometer used, 0,2 g/kg pigment content was chosen as a class interval on the abscissa, and generally 12 to 14 class intervals were used.

[3] American Spice Trade Association

From the data grouped into class intervals, the statistical characteristics $\bar{x}, s, R, V$ were calculated as follows [7]

$$\bar{x} = c + \frac{h}{n} \sum_{i=1}^{k} f_i d_i \qquad (1)$$

$$s = \pm \sqrt{\frac{h^2}{h-1} \left[\sum_{i=1}^{k} f_i d_i^2 - \frac{1}{n} \sum_{i=1}^{k} (f_i d_i)^2 \right]} \qquad (2)$$

$$V = \frac{s}{x} \cdot 100 \qquad (3)$$

$$R = X_{\max} - X_{\min} \qquad (4)$$

where:

- c = midpoint of the class interval of greatest frequency
- h = width of a class interval ($=0{,}2$ g/kg pigment content)
- f_i = frequency in the i-th class intervals
- k = number of the class intervals
- x_i = the midpoint of the i-th class intervals

$$d_i = \frac{x_i - c}{h}, \text{ and}$$

$$n = \sum_{i=1}^{h} f_i$$

The results of the calculations are shown in Table I. The 42 histograms are not presented here owing to lack of space, but their evaluation is discussed below.

After estimating the statistical characteristics, the variance analysis should have to been carried out from two points of view, namely according to qualities, and to years. Considering that for the year 1959 data were not available for each quality, the variance analysis has been carried out for 8 years and for 5 qualities.

Table I

Statistic Characteristics of the Pigment Contents of the Milling Products of Paprika in the years 1959—1967

Years	Exquisite-delicacy					Noble-sweet					Half-sweet					Rose					Hot				
	n	$\bar{x}$	$\pm s$	R	V%	n	$\bar{x}$	$\pm s$	R	V%	n	$\bar{x}$	$\pm s$	R	V%	n	$\bar{x}$	$\pm s$	R	V%	n	$\bar{x}$	$\pm s$	R	V%
1959	58	2,93	0,69	1,97	23,5	104	2,75	0,50	2,44	18,1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1960	131	2,92	0,36	1,92	13,0	243	2,65	0,29	1,61	10,9	52	2,41	0,29	1,72	12,0	52	2,04	0,45	1,39	22,1	9	1,39	0,79	1,28	56,8
1961	136	2,96	0,45	2,12	15,2	253	2,62	0,38	2,17	14,5	59	2,36	0,37	1,73	15,6	58	1,89	0,48	1,96	25,4	30	1,54	0,36	1,26	23,4
1962	166	3,45	0,45	2,47	13,0	275	3,19	0,31	1,69	9,7	98	2,51	0,48	2,45	19,1	75	1,96	0,48	1,85	24,5	35	1,29	0,26	1,02	20,2
1963	165	3,35	0,48	2,35	14,3	314	2,98	0,33	1,71	11,1	123	3,00	0,37	2,09	12,3	79	2,50	0,35	1,74	14,0	45	1,95	0,41	1,94	21,0
1964	231	3,31	0,42	1,71	12,7	308	3,01	0,37	1,91	12,3	171	2,55	0,37	2,96	14,5	60	2,51	0,38	1,91	15,1	33	1,92	0,42	1,65	21,9
1965	435	2,94	0,53	2,96	18,0	392	2,52	0,31	2,30	12,3	392	2,49	0,31	2,30	12,4	78	2,06	0,41	1,56	19,9	37	1,96	0,40	2,55	20,4
1966	208	3,03	0,48	3,21	15,8	335	2,51	0,38	2,23	15,1	596	2,44	0,53	2,89	21,7	120	2,11	0,45	1,83	22,2	50	1,81	0,46	2,28	25,4
1967	335	3,62	0,51	2,54	14,1	282	2,76	0,32	2,44	12,8	286	2,13	0,58	2,95	27,2	115	2,43	0,44	2,15	18,1	76	2,50	0,43	1,77	17,2

n = number of the samples studied
$\bar{x}$ = mean of total pigment content in g/kg expressed in capsantin, according to Benedek
$\pm s$ = standard deviation g/kg
R = range
V = relative deviation

According to literature (8), it must be ascertained before carrying out the variance analysis whether the samples to be compared can be supposed to have been derived from a fundamental set of the same variance. This can be established with the method known as BARTLETT's test.

According to BARTLETT's test, if the value of χ^2 to be found the table belonging to the corresponding probability is not less than that calculated from equation (5), then the fundamental assumption of the variances being equal will hold, otherwise a significant deviation is pointed out by the test, and thus it can be established that the samples are not derived from sets of normal distribution with the same variance.

χ^2 was calculated with the following equation:

$$x^2 = \frac{2.3026}{A}\left[f \log s^2 - \sum_{j-1}^{r}(n_j-1)\log s_j^2\right]$$

where:

r = the number, of groups ($r=8$)

$n_1, n_2 \ldots n_r$ = the number of the samples in the corresponding group

$$f = \sum_{j=1}^{r} n_j - r$$

$$s^2 = \frac{1}{f}\sum_{j=1}^{r}(n_j-1)s_j^2$$

$$A = 1+\frac{1}{3(r-1)}\left[\sum_{j=1}^{r}\frac{1}{n_j-1}-\frac{1}{f}\right]$$

The χ^2 value given in the table for the 7 degrees of freedom of the five examined milling products of different qualities, the calculated χ^2 value, as well as the comparison of these values are to be found in Table II.

Table II

Results of The Bartlett Test

Qualities	Degrees of freedom	χ^2 0,95	χ^2 calculated	Comparison	Discussion
Exquisite-delicacy	7	14,1	28,90	$\chi^2c > \chi^2$ 0,95	Significant deviation
Noble-sweet	7	14,1	18,04	$\chi^2c > \chi^2$ 0,95	Significant deviation
Half-sweet	7	14,1	176,00	$\chi^2c > \chi^2$ 0,95	Significant deviation
Rose	7	14,1	16,41	$\chi^2c > \chi^2$ 0,95	Significant deviation
Hot	7	14,1	28,30	$\chi^2c > \chi^2$ 0,95	Significant deviation

As can be seen, there is a significant deviation among the variances in all cases, therefore the variance analysis had to be given up in order to avoid the rather tedious calculations for fundamental sets of different variances. We have tried instead to use the so-called non parametrical rank-sum test, as proposed by ZUKÁL [9].

Tishmethod does not require either the normal distribution of the random variable nor the identity of the variences of the fundamental set. In KRAMER's rank-sum test [10] rank-numbers are assigned to the mean values, so that rank-number 1 is given to the highest mean, whereas the sequence of natural-numbers is assigned as rank-number to the other mean values arranged according to their decreasing order. The sum of the rank-number has to be evaluated, taking into account the degrees of freedom and the level of probability on the basis of the table given by KRAMER.

Table III

Kramer's Rank-sum Test according to Qualities

Years	Qualities				
	Exquisite-delicacy	Noble-sweet	Half-sweet	Rose	Hot
1960	1	2	3	4	5
1961	1	2	3	4	5
1962	1	2	3	4	5
1963	1	3	2	4	5
1964	1	2	3	4	5
1965	1	2	3	4	5
1966	1	2	3	4	5
1967	1	2	3	5	4
Sum of all ranks	8	17	23	33	39

Table III shows the rank-numbers assigned to the means of the pigment content of different qualities and their sums, Table IV the rank-numbers of the means arranged according to the years.

Table IV

Kramer's Rank-sum Test According to the Years

Qualities	Years							
	1960	1961	1962	1963	1964	1965	1966	1967
Exquisite-delicacy	8	6	2	3	4	7	5	1
Noble-sweet	5	6	1	3	2	7	8	4
Half-sweet	6	7	3	1	2	4	5	8
Rose	6	7	8	2	1	5	4	3
Hot	7	6	8	3	4	2	5	1
Sum of all ranks	32	32	22	12	13	25	27	17

Results and Discussion

The histograms showing the distribution of the results of measurements of the pigment content of different qualities for the same year do not show the typical shape of a normal distribution, though they are based on a rather high number of samples, instead, they are rather of a multimodal type. The deviations from the normal distribution are proved by both the probit calculation and the empirical distribution function yielded by the graphs representing the cumulative frequencies.

The deviation from the normal distribution is probably due to the circumstance that several distributions are involved, which also contributes to the increased values of the standard deviation.

The high values of standard and relative deviations (Table I) can also be explained by the fact that in the method of qualification used at present the colour observed visually plays a determining role. The variance may have been increased by the circumstance that special requirements were to be met at certain markets, *e.g.* when the export paprika product had to be hot and of a large pigment content at the same time. Had it been possible to collect the data of the mentioned samples separately, the variances would have certainly been less.

However, the values of the variance of milling products of the same quality for 8 years can not be considered as the elements of a fundamental set of the same variance on the basis of the results of BARTLETT's test.

The comparison of the means of the different qualities using KRAMER's rank-sum test results, however, in a clear understanding of the facts. It can be seen from Table III that the mean pigment content of the milling products of exquisite-delicacy paprika was the highest in each year. For noble-sweet products there was only one year (1963) in which the mean of the pigment content was less than that of the half-sweet products. The mean of the pigment content of the half-sweet product was higher in every year than that of the rose product; similarly, the rose product shows a higher mean value than the hot milling product, except for the year 1967.

According to the table given by KRAMER, if the sum of the rank-numbers in the table 8×5 is between 17 and 31, then there is no significant difference. It can be seen from Table III that the sum of the rank-numbers according to qualities shows a significant difference for exquisite delicacy, noble-sweet, rose and hot milling products.

In Table IV, arranged according to years to reveal the influence of the crop-years, the sum of the rank-numbers should be beyond the range 10 to 35 for the difference to be significant. This condition not being fulfilled in either of the columns of table IV, significant differences among the years cannot be found.

On the basis of the above, the conclusion can be drawn that the qualification carried out chiefly on the basis of colour standard samples can be suitably replaced by a qualification based on the pigment content; whereas the fluctuations found in the pigment content of the same quality for different years are not significant. From this it can be concluded that the preparation of annual standard samples is not indespensable.

The real values of the pigment content of milling products qualified with the method of characteristic samples can be given as follows.

Pigment content of the product, expressed in capsantin:

exquisite products:	$3{,}2 \pm 0{,}4$ g/kg,
noble-sweet:	$2{,}8 \pm 0{,}4$ g/kg,
half-sweet:	$2{,}4 \pm 0{,}4$ g/kg,
rose:	$2{,}1 \pm 0{,}4$ g/kg,
hot:	$1{,}7 \pm 0{,}4$ g/kg.

It can be proposed that the values of the total pigment content should be considered as a basis of objective qualification, instead of the colour standard samples used at present.

The qualification carried out on the basis of the total pigment content gives a more exact expression of the real value of the milling product; the parameter can be constantly followed up during the processing and furthermore, by its use it is possible to apply linear programming, for assorting the half-products and determining the quantities to be compounded in order to optimize the technology of production.

* * *

Thanks are due to Dr. T. SZÉLL for his interest and for reading the manuscript. The author is further grateful to Dr. E. ZUKÁL, chief research officer, for his advices and for the kind aid given in statistical analysis.

References

[1] *Huszka, T.:* Konzerv és Paprikaipar (Canned Food and Paprika Industry) **4**, 109 (1968)
[2] Hungarian Standard MSZ 11851—62.
[3] *Benedek, L.:* Z. U. L. (Journal of Food Control and Research) **107**, 228 (1958)
[4] *Palotás, J., I. Konecsni:* ÉVIKE (Food Inspection Bulletin) **10**, 200 (1964).
[5] ASTA Analytical Methods, Method 19., p. 44—45.
[6] *Cholnoky, L.: L. Zechmeister:* M. Chem. F. (Hungarian Journal of Chemistry) **32**, 97 (1926)
[7] *Vincze, I.:* Statisztikai minőségellenőrzés (Statistical Quality Control), Közgazd. és Jogi Kiadó, Bp. 1958), p. 25—34.
[8] Ibid., p. 102—105.
[9] *Zukál, E.:* Personal communication.
[10] *Kramer, A.:* Food Techn. **14**, 576 (1960).

МАТЕМАТИЧЕСКО-СТАТИСТИЧЕСКАЯ ПРОВЕРКА АНАЛИЗОВ СОДЕРЖАНИЯ КРАСЯЩИХ ВЕЩЕСТВ В ПОМОЛЬНЫХ ПРОДУКТАХ СЕГЕДСКОГО КРАСНОГО ПЕРЦА ЗА 1959—1967 ГГ.

Т. Хуска.

Автор проверяет изменения содержания всех красящих веществ, являющихся одной из важнейших качественных характеристик в помольных продуктах красного перца в 1959—1967 гг. За изучаемый период на основе результатов 7100 анализов пяти разных сортов помольных продуктов он устанавливает обычные статистические характеристики ($\bar{x}$, s, R, V). На базе „непараметрической" пробы *Крамера* он обнаруживает сигнификантное различие содержания красящих веществ в помольных продуктах разного качества за проверяемый период, в то же время он подчёркивает, что различия того же сорта за разные годы не являются сигнификантными.